DOCUMENT RESUME

ED 412 223 TM 027 423

AUTHOR Cox, Philip; Godfrey, John R.
TITLE The Importance of Assessment Procedures to Student Learning Outcomes in Religious Education.
PUB DATE 1997-03-00
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Catholic Schools; Educational Assessment; *Educational Testing; Feedback; Foreign Countries; *High School Students; High Schools; Knowledge Level; *Outcomes of Education; Pretests Posttests; *Religious Education; *Test Use; Urban Schools
IDENTIFIERS Australia

ABSTRACT

In Perth, Western Australia, summative assessment has not been a teaching tool in the teaching of religious education courses in the Catholic schools. This study investigated whether the use of formal assessment procedures in the teaching of religion had an effect on student learning outcomes. Subjects were 128 students (4 classes) in year 8 of an urban Catholic high school. The individual class variation in scores was nested in the variation of scores between the experimental and control groups. A multiple choice test was given before and after instruction to measure student knowledge. Students in the experimental group were quizzed on work covered in each teaching module, given feedback from the testing, and motivated to prepare thoroughly for the final test. Posttest results indicate that scores of the experimental group were higher than those of the control (untested) classes. Treatment given the experimental classes does seem to have resulted in significant differences in learning outcomes. Results support the view of G. Rossiter (1981) that a relationship exists between clarity of purpose and learning outcomes. Appendixes present summaries of the posttest results. (Contains two tables, two figures, and seven references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 412 223

The Importance of Assessment Procedures to Student Learning Outcomes in Religious Education.

A paper presented to the

American Educational Research Association 1997 Annual Meeting

held in

Chicago, March 24 to March 28, 1997

by

Philip Cox and John R. Godfrey

of

Edith Cowan University
Churchlands
Western Australia

TM027423

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

John Godfrey

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

THE IMPORTANCE OF ASSESSMENT PROCEDURES TO STUDENT LEARNING OUTCOMES IN RELIGIOUS EDUCATION

BY

PHILIP COX

AND

JOHN GODFREY

EDITH COWAN UNIVERSITY

CHURCHLANDS

Assessment and Evaluation: Aspects of Teaching

The process of utilising assessment and evaluation within the context of education relates to the principles of good teaching and classroom management. Bloom, Hastings and Madaus (1971) point out that "one cannot see 'understanding' or observe 'critical thinking'" (p. 33) and so it is necessary for the purposes of meaningful evaluation to develop objectives stated in terms of "more readily observable outcomes or changes on the student's part" (p. 22). This phase of the teaching process is necessary because educational objectives are often very broad in their scope and as such are often vague and hence "cannot serve as an instruction or educational model" (p. 21). The teacher must therefore interpret these broad objectives and establish specific and tangible objectives. This step enables the teacher to discover if aspects of the subject have been taught. This element ties this stage of the teaching process into evaluation and assessment.

The importance of measurement, assessment and evaluation techniques to the teaching process relates to the reason for the process of teaching itself. One assumes that students will be different after a unit of work has been taught. The question arises as to the degree of difference. Hence measurement, assessment and evaluation are important to determine the degree of difference. Within this context, the main purpose of classroom instruction is to enable students to achieve intended learning outcomes. In so doing the teacher becomes a predictor. The teacher needs to decide to utilise a particular technique "'X' rather than 'Y' because it is predicted that 'X' will be more effective in producing a desired outcome in the learner" (Lee, 1973, p. 41). This requires evaluation of the technique chosen and thus the need for assessment arises. The teaching process requires that assessment and evaluation occur. In this way assessment is not a post teaching procedure, it is an integral part of the teaching process.

Cole and Chan (1987) are particularly wary of teachers who are overtly confident of their capacities to make informal judgements about a student's abilities and achievements. They classify this type of teacher as a 'self-reliant assessor' (p. 295). They point out that teachers who shy away from assessment and evaluation strategies on some philosophical ground or principle, believe that they can answer the questions relating to effective teaching without utilising the vast wealth of objective information that can be gained through the use of effective diagnostic, formative and summative evaluation.

Background

The teaching of religious education in Perth, Western Australia has in the past not utilised summative assessment procedures as a teaching tool. Students therefore had no experience of testing in religious education. A great deal of debate within the literature and at the classroom level centeres around the issue of using assessment procedures in the religious education classroom. It seems that many teachers feel that the teaching of religious education is somehow different from the teaching of other subjects and hence should not (or could not) utilise assessment procedures. This lack of exposure to testing in religious education classes created an ideal situation in which to set up an experiment to ascertain the importance of assessment procedures to student learning.

Aims of the study

The aim of this study is to investigate whether the use of formal assessment procedures in the teaching of religious education has an affect on student learning outcomes.

Subjects

The subjects were 160 students in Year 8 (the students eighth year of formal education) in a metropolitan Catholic High School in Perth, Western Australia. Apart from ensuring a gender balance the students had been randomly allocated to each class.

Initially eight religious education teachers were involved in the study. Four classes were randomly selected to represent the experimental group. One of the four control class teachers withdrew support for the study part way through the experiment leaving only three classes to represent the control group. Given that 77 students, from 3 separate classes remained in the study the loss of one class was not seen as detrimental to the outcomes of the study. The experimental group contained four classes totalling 128 students.

Design

A nested experimental design was utilised to provide the necessary data and to draw conclusions to answer the research questions. There are two levels of effect within the study. The individual class variation in scores is 'nested' within the variation of scores between the experimental and the control groups. Factor A is treatment / non-treatment and represents the first level of analysis. At this level the two groups include the Experimental Group and the Control Group. The Experimental Group experienced a range of formal assessment procedures (treatment). The Control Group did not experience this treatment. Factor B, at level 2, separates the experimental and control groups intotheir individual classes. At this level variation of test scores between individual classes is the focus of the analysis. The experimental design is shown in Table 1.

Table 1

Nested design of the study

Level 1	Experimental Group	Control Group
Factor A	(Treatment - Formal assessment procedures)	(Non-Treatment - No formal assessment)
Level 2	Class 1-4	Class 5-7
Factor B	(Teacher differences)	(Teacher differences)

Knowledge tests

To ensure consistency of scoring of the knowledge test it was decided that a twenty item four choice multiple choice test would be used. Through a series of pilot studies in other schools the test items were gradually refined to produce effective distracters. While in some items more than 25% of the students scored the correct response the average item difficulty for this group remained very near 25%. This is well within the range of 20% to 80% set by Kubiszyn and Borich (1987, p. 29).

Reliability and validity of the knowledge test

The knowledge test was found to be reliable and valid. Internal consistency was tested using a split half reliability index. An odd-even split-half reliability index of .82 was obtained for the knowledge test.

A discrimination index for each of the twenty knowledge questions was also calculated to indicate the reliability of individual items. To determine this index the upper and lower group boundaries were set at 27%. The average discrimination index is .34.

The post test and follow-up test design of this study enabled a calculation of a stability reliability index. Given that there was no intervention between these two tests the reliability index was calculated using these two tests. The time span between the post test and the follow-up test tests was two weeks. This analysis produced a Pearson r of .87 indicating a high degree of similarity between the scores on each test occasion. This result indicates that the knowledge test is reliable.

Validity of the knowledge test was indicated through content validity. This process ensures that the items of the knowledge test are drawn from the domain of objectives set out in the module. Each objective is represented by one item in the knowledge test. The test items were selected to ensure that no aspect of the unit was over represented in the tests.

Procedure

The teachers in the experimental group were intensively inserviced on the methodology of teaching that was required to ensure uniformity of treatment in the four experimental classes. This inservicing explained that the treatment to be given to the experimental group was to involve the use of formative and summative assessment. The treatment would involve revising previous lessons, setting homework and home study. Students would be quizzed on work covered during

the module, given feedback in each subsequent lesson and frequently motivated to prepare thoroughly for the final test. Normally this approach to teaching has not been part of the methodology of teaching religious education in Catholic schools in Western Australia. The control groups would not receive this treatment nor would the teachers in the control group have this information. Observation and recording of teaching in the control group is used to confirm the level of use of systematic assessment procedures.

Each teacher in the experimental group was given a teaching programme and daily lesson plans. The lesson plan included review questions, homework and class work. In an effort to prevent teachers teaching to the tests none of the teachers had access to test papers until the morning designated for each particular test. The daily review tests were administered to the experimental group, were collected and marked by the researcher and returned prior to the next lesson. The teachers then went through each item, corrected any misunderstandings and directed students to correct errors or incomplete answers. All classes were given a pretest prior to the commencement of the study. All classes were given the same test as a post test at the end of the four week module. Two weeks later, after two weeks of holidays, a follow-up test was administered.

The analysis of the knowledge test scores utilised the procedures outlined by Dayton (1970) for a nested design with unequal class sizes. An additional complication arose due to the unequal number of classes in each group. To eliminate this complication the mean scores of the three control classes was

averaged and then multiplied by four. Through this process, the mean results for the experimental group (four classes) could be compared with the mean score of the control group (three classes).

Results

Table 2 summarises the scores of the knowledge tests. The knowledge pretest scores indicate that no one class has a score in the knowledge pretest that is markedly different from any other class. The mean score on the knowledge pretest for each class also indicated that no significant knowledge of the content of the unit existed. The sample mean was 5.14 with a standard deviation of 1.93.

Table 2

Mean test scores for each class.

Class	Pretest Mean Score	Posttest Mean Score	Follow-up Test Mean Score
1	6.0	13.5	12.2
2	5.2	9.9	9.3
3	4.9	12.6	10.7
4	4.8	11.2	10.0
5	4.7	5.9	5.4
6	5.0	5.6	5.2
7	5.3	4.9	5.6

Classes 1 - 4 are the Experimental Classes; Classes 5 - 7 are the Control Classes.

Each individual class had similar results with a similar distribution. The mean scores of the experimental and control groups were also very similar, 5.22 and 5.05 respectively.

The difference between the experimental and the control groups, when the knowledge pretest scores are considered, is not significant at the 0.05 level $t(158) = 0.54$, $p > 0.05$. An ANOVA of the results of the seven classes indicates that no two classes are significantly different at the 0.05 level $F(6, 153) = 1.19$, $p > 0.05$.

Tests for skewness indicated that the knowledge and values scores did not differ significantly from the normal distribution at the pretest, post test or at the follow-up test stage.

The results of the post test illustrate that a difference exists between the experimental and control classes. Each of the experimental classes scored mean post test results well above the means of the control classes. The experimental classes had means of 13.5, 9.9, 12.6 and 11.2 while the three control classes had mean scores of 5.9, 5.6 and 4.9. The standard deviation of each class was very similar ranging from 2.4 to 3.2. The experimental group had an average score 11.9 while the control group had an average score of 5.4 The standard deviation of the scores of the experimental group was 3.3 while the control group had a standard deviation of 2.5.

The change in scores between the pretest and post test scores also indicates that the control and the experimental classes were very different. The four experimental classes improved the mean score by 7.5, 4.7, 7.8 and 6.4. The standard deviations were 3.5, 2.7, 4.0 and 3.0 respectively. This is in contrast with

the three control classes where the mean score showed very little change. The means changed by 1.1, -0.2 and by 0.2. The standard deviations were 2.9, 2.5 and 2.8 respectively. To further illustrate the difference between the control and the experimental groups the mean difference between the pretest and the post test score for the control group was 0.4 and 6.7 for the experimental group.

The results of the knowledge post test indicate differences at both levels of the nested design. The three control classes have shown almost no change in score. Figure 1 gives a visual impression of the degree of change that occurred between the knowledge pretest and knowledge post test. It shows that each of the four experimental classes had scores that improved after the pretest. The small amount of change in the scores of the control classes is also very evident. Figure 2 illustrates the change in scores for the experimental and the control groups and again shows the difference between the results.

Further analysis of these results confirms the impressions evident in Figure 1 and Figure 2. This result indicates a significant level of difference in test scores at the two levels of the nested design. The nested design analysis indicates that the variation in post test knowledge scores is significantly different at the 0.05 level when 'method' is considered (Appendix A). The differences between individual teachers was not significant at the 0.05 level.

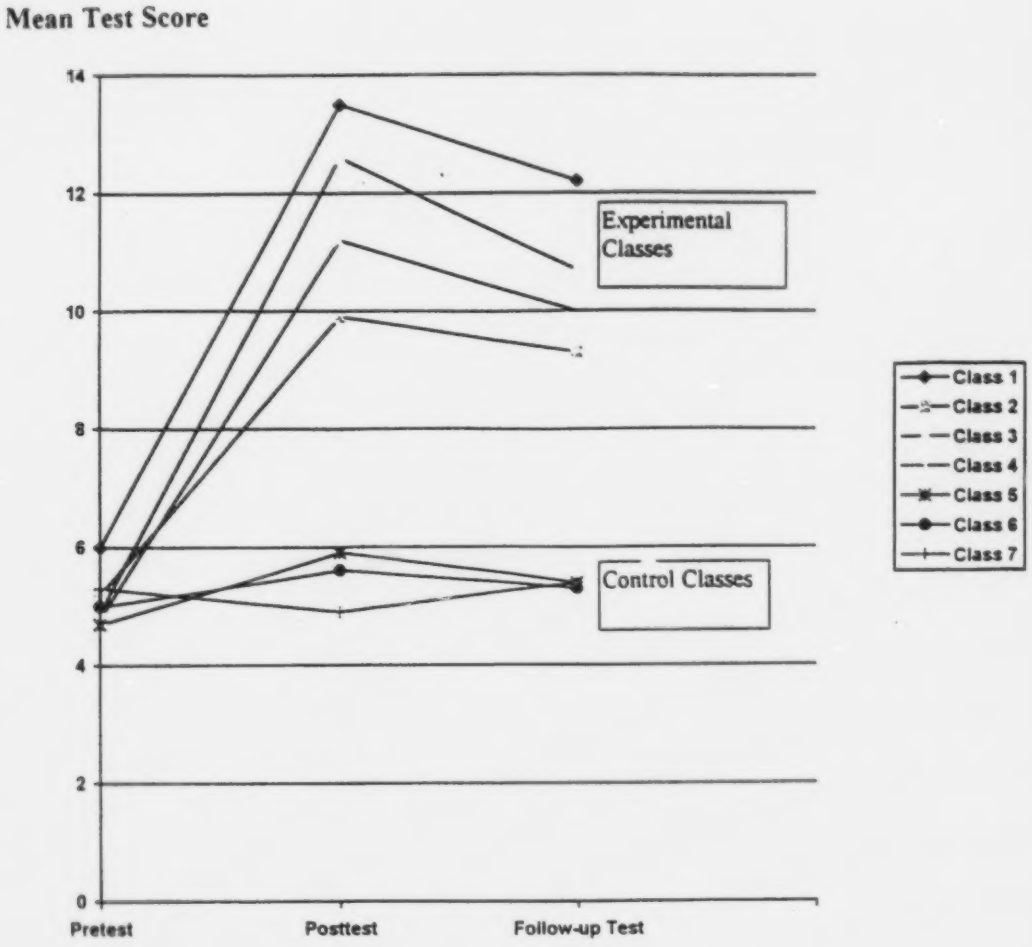

Figure 1. Mean knowledge test scores for individual classes.

The treatment given to the experimental classes does seem to have resulted in significant differences in knowledge learning outcomes. These differences exist when the results of each class are compared and when the individual experimental

BEST COPY AVAILABLE

and control class results are combined to form two groups. The knowledge test results indicate that significant differences exist between the experimental and control groups.

The results of the follow-up test indicate that the four classes representing the experimental group scored at a higher level than the three classes in the control group.

The mean score for the experimental classes was 10.65 with the mean scores of the four experimental classes ranging from 9.33 to 12.22. The mean score for the control classes was 5.44.

The mean scores of the control classes are little different from the pretest scores. The ANOVA indicates that there is no significant difference, at the 0.05 level, between the pretest and post test scores (Appendix A). The mean change in test score between the pretest and the follow-up test for each of the three control classes was generally less than 1 point. The mean change for the control group was 0.39. The level of change for the four experimental classes was more substantial. The four classes recorded mean changes of 6.2, 4.2, 5.8 and 5.3. The average change in score for the experimental classes was 5.4. The standard deviation for each class was very similar ranging from 2.2 to 3.8. The standard deviation for the control group was 2.4 while the experimental group had a standard deviation of 3.8.

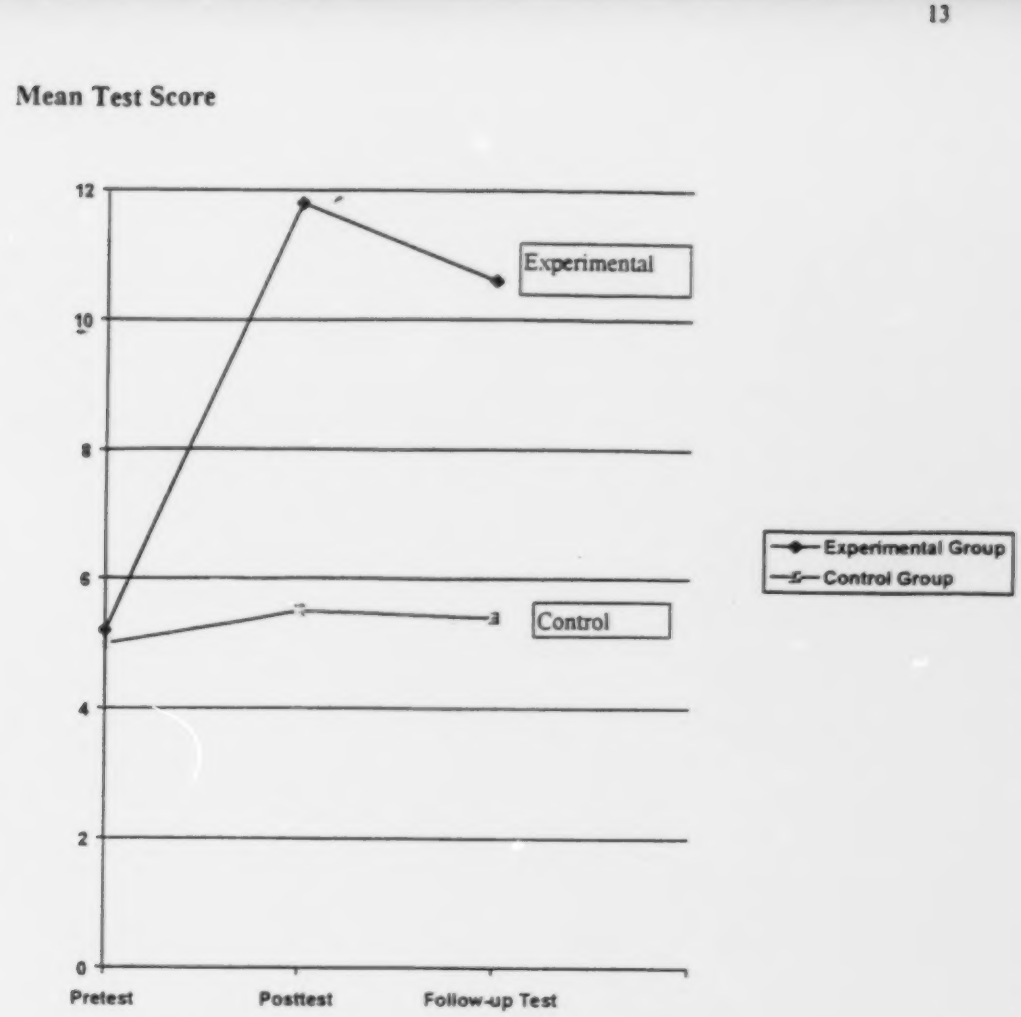

Figure 2. Mean knowledge test scores for the experimental and the control groups.

Figure 1 gives a visual impression of the degree of change that occurred between the knowledge pretest scores and follow-up test results. It shows that each of the four experimental classes had scores that maintained their level at the follow-up test stage. The amount of change in the scores of the control classes remains at a very low level indicating little change from the pretest results. Figure 2 illustrates the change in scores for experimental and the control groups and again

BEST COPY AVAILABLE

shows the difference between the results of these two groups. The experimental group maintained a significantly higher score in the follow-up test despite the intervening two week period.

The knowledge test results indicate significant differences between the experimental and control groups. The nested design analysis (Appendix B) indicates that the variation in the follow-up test knowledge scores is significantly different at the 0.05 level when `method' is considered. The differences between individual teachers was not significant at the 0.05 level.

Discussion and Implication

Significant differences between the control and the experimental groups were evident when the knowledge test scores were analysed. The pretest scores indicated that all students had similar levels of knowledge prior to teaching. At the post test stage the control group had shown little change in test scores while the experimental group illustrated significant change in test scores. The difference between the two groups was maintained at the follow-up test stage.

The results clearly supported Rossiter's (1981) view that a relationship exists between clarity of purpose and learning outcomes. The students who received the `treatment' had results that were significantly better than the in the control group.

The nested design of this study allowed individual classes to be compared, as well as a comparison of the experimental group and the control group. In both instances the results of the experimental classes were significantly different to the results of the control classes. The results of each control class were similar, and indicated that little learning of content had occurred. The results of the four

experimental classes were similar to each other and indicated a significant positive change in knowledge test scores between the pretest and the post test. These results therefore seem to support Rossiter's contention that clarity of purpose can directly influence learning outcomes. The four experimental classes had direction and purpose. The control classes did not have this level of clarity.

Content that had been covered by the teachers of the three control classes seemed to have not been learnt. Post test and follow-up knowledge test results indicated almost no change in knowledge test scores from the scores attained by the students prior to the module beginning. Teachers were 'teaching' but the content was not being learnt. This result was in contrast to the observed outcomes of the experimental classes. Here, teachers imposed a formal assessment structure, actively revised each lesson, set minor tests, reviewed material and actively utilised many forms of formative assessment. In these classes students learnt the material that was being taught. Knowledge post test scores were significantly higher than the pretest scores. Learning was shown to be long term as the follow-up test results were also significantly higher than the pretest scores and test scores were maintained after the post test.

The differences between the experimental and the control post test and follow-up knowledge test scores cannot be explained by differences that existed between the classes prior to the study beginning. This has been shown with an analysis of a broad range of indicator variables. These included indicators of religious background, commitment to religion, home study, prior knowledge of the unit of

work and reading ability. The observed differences in knowledge test scores must therefore be associated with the treatment the experimental classes received during the study.

In calculating the within group difference as well as the between group differences, the nested design analysis allows comment on the possible differences between each teacher in the control and experimental classes. While every care was taken in the experimental design to randomly allocate teachers to each class, some advantage could have occurred for the experimental classes. These teachers may have been more dynamic, more committed and more inspirational. The nested design analysis indicated that when the scores of individual classes were compared there were no significant differences. This pattern was evident for the knowledge test at the pretest, post test and follow-up test stages. The nested design analysis indicated that there was no significant difference between any of the four experimental classes when the post test and follow-up test results were considered. Similarly the analysis indicated that there was also no significant difference between any of the three control classes. This indicates that teacher differences in this study did not significantly influence the knowledge test scores. It would seem that the difference in test scores was the result of the difference in teaching.

A significant theme in the literature pointed to the effect of poor teaching within religious education in Catholic schools. This perception was shown to hold true within the study school. Observation of the control group of classes indicated that the teaching lacked academic rigour. No tests were planned, teachers failed to utilise any structured formative or summative assessment procedures. In these

three classes knowledge test scores were very low. Scores at the end of a four week module were barely different from the scores recorded in the pretest. No learning appeared to have taken place. On the other hand the four experimental classes showed significant changes in knowledge test scores. Teaching in these classes included systematic formative and summative assessment. They were shown to do much more study. It would seem that the concern expressed in the literature regarding teaching technique in religious education is supported by the results of this study.

The problems facing religious education in Catholic schools have been viewed too exclusively as problems of 'religion' rather than problems of education. The literature faces this issue from an educational perspective. The literature calls for a more professional approach to the teaching of religious education. This professional approach involves determining objectives, determining classroom process and designing methods for determining whether the classroom processes achieves the objectives. Thus the need for assessment and evaluation is integral for good education. As good education is integral for religious education, the inclusion of assessment and evaluation is crucial for a professional approach to teaching religious education in Catholic schools. The results of this study confirm that the use of assessment and evaluation in the teaching of religious education is of benefit to both the student and the teacher.

The students in the classes who were told about the final test performed at a significantly higher level than those who had no knowledge of this end of module test. The focus of this long term goal was maintained with daily tests. Students

knew that each day their learning would be tested and their results constantly reviewed. Students quickly see the direct connection between the effectiveness of their home study and the results of their daily tests.

Conclusions

At level one of the nested study, a clear difference between the experimental group and the control group is observed. This difference was evident not only at the post test stage but continued beyond the teaching phase and was evident in the follow-up test. These results indicate that the treatment was able to produce significant change in knowledge learning outcomes. The treatment involved the use of assessment and evaluation procedures in the teaching of religious education. The control group was not exposed to this method of teaching. The results of the control group indicated that no significant change in knowledge learning outcomes occurred between the pretest, posttest and at the follow-up test stage.

Analysis of a range of indicator variables which may have an influence on student learning indicated that there was no significant difference between the profile of the control and the experimental groups. Relating knowledge test scores to these variables indicated no significant relationship. Knowledge test scores did not significantly vary when each factor was considered. A student's religious background and commitment to the Catholic religion did not appear to impact on knowledge learning outcomes.

The elimination of each of these extraneous variables leaves the 'treatment' as an intervening variable on student learning outcomes. The differences in knowledge learning outcomes can therefore only be accounted for by the difference in teaching methodology.

At level two of this nested design the conclusions are the same. Level two considered individual class differences. The analysis of knowledge results indicated that while small differences in knowledge scores were evident between each of the four experimental classes these differences were not significant. This was the case at all three stages of testing. The same outcome arose when the knowledge scores of the three control classes were compared. Individual teacher differences therefore did not complicate student learning outcomes in this study.

Each of the four experimental classes scored significantly higher knowledge test results than each of the three control classes. The extraneous variables (religious background, commitment to the Catholic religion) were also considered at level two of this analysis. No differences were evident indicating that all classes had similar personal and family characteristics. These factors were shown to not have any significant effect on student learning outcomes.

The results of the study are clear. The use of a more academic mode of teaching, with its associated assessment and evaluation procedures, in religious education in Catholic schools does affect the knowledge learning outcomes of students. The learning effect is significant and positive. The students who did not receive the treatment indicated little change of knowledge scores. The students

who did receive the treatment demonstrated significant gain in knowledge scores. Therefore change in knowledge scores was not the result of other factors but may be directly attributable to the teaching process.

Some teachers of religious education believe that their subject is different from subjects such as mathematics, science and history. They believe they can teach effectively without the benefits of assessment and evaluation. It is important to consider the results of this study in the light of incorporating assessment and evaluation procedures in the teaching methodology of religious education.

References

Bloom, B., Hastings, J. T. & Madaus, G. F. (1971). Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill.

Cole, P. G., & Chan, L. K. S. (1987). Teaching principles and practice. Sydney: Prentice Hall.

Crawford, M. L., & Rossiter, G. (1989, February). Religion teachers as missionaries to a teenage culture. Word in Life, pp. 17-20.

Dayton, C. (1970) . The design of educational experiments. New York: McGraw-Hill.

Kubiszyn, T., & Borich, G. (1987). Educational testing and measurement. Classroom application and practice. Glenview, Illinois, USA. Scott, Foresman and Company

Lee, J. M. (1973). The flow of religious education: A social science approach. Birmingham, Alabama: Religious Education Press.

Rossiter, G. (1981). The gap between aims and practice in religious education. Word in Life, 29, 24-28.

Appendix A

Summary of nested design analysis of knowledge posttest results

Source	df	S of S	MS
Methods(A)	1	1367.65	1367.65
Teachers B(A)	6	183.42	30.57
Error	25	1178.50	47.14

	F Ratio
Method $\frac{MS(A)}{MSB(A)}$	44.74*
Teacher $\frac{MS_{BA}}{MS_{error}}$	0.65**

Appendix B

Summary of nested design analysis results of knowledge follow-up test results

Source	df	S of S	MS
Methods(A)	1	876.60	876.60
Teachers B(A)	6	103.95	17.33
Error	25	1128.83	45.15

	F Ratio
Method $\frac{MS(A)}{MSB(A)}$	50.58*
Teacher $\frac{MS_{BA}}{MS_{error}}$	0.38**

Note: * Significant at the 0.05 level ** Not significant at the 0.05 level